

A New Perceptual Quality Measure for Bit Rate Reduced Audio

Thilo Thiede, Ernst Kabout

*Technical University of Berlin, Institute for Telecommunications¹,
Berlin, Germany*

For the quality evaluation of perceptual audio codecs, appropriate measurement algorithms are needed, which detect and assess audible artefacts by comparing the output of the codec with the uncoded reference. A filter bank based perceptual model is presented, which yields better temporal resolution than FFT-based approaches and thus allows a more precise modelling of pre- and post-masking and a refined analysis of the envelopes within each filter channel.

0. Introduction

Perceptual audio coding algorithms perform a drastic irrelevancy reduction in order to achieve a high coding gain. Signal components that are assumed to be unperceivable are not transmitted and the coding noise is spectrally shaped according to the masking threshold of the audio signal. Simple quality measures (e.g. signal to noise ratio, harmonic distortions), which can not separate these inaudible artefacts from audible errors, can not be used to assess the performance of such coders. Since subjective listening tests are very time consuming and expensive, there is a strong demand for new measurement systems that are capable to estimate the perceived audio quality of such perceptual coders. Those perceptual measurement methods detect and assess audible artefacts by comparing the output of the codec with the uncoded reference. Most of the known approaches are based on perceptual models for steady state signals. In order to limit computational complexity, they use an FFT to perform the spectral decomposition of the input signals. To achieve a sufficient spectral resolution even in the lowest auditory filters, block lengths of 1024 or 2048 samples are used. The resulting time window of 20 to 40 ms is short enough to model post-masking and temporal integration, but could be too long to model pre-masking, which in some cases lasts only a few milliseconds [1]. Filter bank based approaches yield much better temporal resolution and are thus allowing a more precise modelling of pre-masking. The temporal fine structure of the envelopes at each auditory filter is preserved and can be used to achieve additional information about the signals, which may contribute to the perceptibility of coding artefacts.

¹ Both authors work within a research project supported by the Deutsche Telekom AG.

We are introducing a perceptual model that is based on a new auditory filter bank. The centre frequencies of the individual filters are equally distributed over a perceptual pitch scale. The top of the filter shape is slightly rounded to ensure that the chosen number of filters covers the full frequency range without ripples in the overall frequency response. Alternatively, a flat top or a more rounded top (roex-filter) can be chosen. In order to model masking thresholds, the filter slopes decrease exponentially over the Bark scale. The steepness of the slopes can be chosen either at fixed values or level dependent. The filter bank algorithm is rather fast as compared to conventional approaches but still much more time consuming than FFT-like methods.

Out of the filtered representations of the test signal and the uncoded reference, numerous output parameters are calculated, including a reduced noise loudness, envelope correlations and measures for binaural effects. Different configurations of the filter bank were tested. We varied the number of filters (between 50 and 180), the filter shapes (rounded exponentials or triangular), the slopes of the filters (12 - 31 dB/Bark, fixed and level dependent) and the temporal resolution. The results of the model are compared to those of the numerous listening tests performed by ITU and MPEG for the evaluation of perceptual codecs. For most of the available databases, the correlations between model predictions and subjective scores are very high.

1. Principles of Perceptual Measurement

Perceptual measurement methods detect and assess audible artefacts by comparing the output of the codec with the uncoded reference (*Fig. 1*). The input signals are transformed into short time spectra, which are fed into a perceptual model (*Fig. 2*). After weighting the spectral components with the transfer function between outer ear and inner ear, the time-frequency representation of the signal is transformed into a time-pitch representation by grouping neighbouring frequency bands into fixed fractions of critical bands. The time-pitch representations are smeared out over time and frequency in order to model simultaneous and temporal masking. There are two different concepts how to achieve a measure for audible distortions from the perceptual model. The most straightforward approach is to apply the time-frequency smearing only to the reference signal and use it as a masking threshold for the error signal („masked threshold concept“, *Fig. 2a*). The other approach is to use the complete perceptual model on both the coded signal and the reference signal and compare the internal representations of both signals („comparison in the basilar domain“, *Fig. 2b*). Output parameters are either the ratio between error signal and masked threshold [1], the number of blocks in which audible distortions occurred [1], an estimate for the loudness or annoyance of the distortions [2] or a probability for the detection of the distortion by a listener [4][5].

1.1. Possible Problems with FFT-Based Models

The Perceptual models described above use an FFT for the time-frequency decomposition. As the relation between the linear frequency scale resulting from the FFT and the pitch scale needed for perceptual measurement is highly non-linear, there is always a trade-off between temporal and spectral resolution: grouping frequency components into equidistant fractions of critical bands requires a large number of spectral components and thus a rather long time window. On the other hand, a long time window limits the accuracy when modelling pre- and post-masking. Additionally, it will result in a partial loss of information contained in the temporal envelopes within each auditory filter. Such information is necessary when modelling

binaural effects, but can also be used to model dependencies of masked thresholds from the type of masker.

2. Description of the New Measurement System

One task of the new model is to check the influence of different spectral and temporal resolutions on the performance of the model as a predictor of audio quality. This is not possible with an FFT-based model because any refinement of the temporal resolution would affect the spectral resolution in the lower frequency bands and vice versa.

The model uses a new approach to estimate the reduction of coding noise by the masking reference signal. It makes use of the envelope modulations at each auditory filter in order to determine masking thresholds. This allows modelling additivity of masking as well as the reduced masking effect of pure tones as compared to narrow band noise. We did already spend some work in modelling binaural effects but up to now it did not yield significant improvements in the prediction of available listening test results.

2.1. The Auditory Model

Like other perceptual measurement systems, the model estimates the amount of audible distortions by comparing the test signal with an undistorted reference. After filtering the input signals with the transfer function from outer ear to inner ear (Eq. 1, from Terhardt [9]), the signals are fed into the auditory filter bank.

$$A(f) = 10^{-\frac{1}{20} \left[-6.5 \exp \left[-0.6 \left(\frac{f}{\text{kHz}} - 3.3 \right)^2 \right] + \frac{1}{1000} \left(\frac{f}{\text{kHz}} \right)^4 \right]} \quad (1)$$

2.1.1. The Filter Bank

The filter bank consists of an arbitrary number of filter pairs that are equally distributed over a perceptual pitch scale. Therefore, no additional scale transformation is required. Similar to the filter characteristics of an FFT, each filter pair consists of one filter representing the real part of the signal and one filter representing the imaginary part. In the beginning, the envelopes of the pulse responses of each filter have a raised cosine shape:

$$h_{re}(t) = \cos^2(\pi \cdot bw \cdot t) \cdot \cos(2\pi \cdot f_c \cdot t) \quad \left| \begin{array}{l} |t| < \frac{1}{2 \cdot bw} \\ bw: \text{bandwidth} \\ f_c: \text{centre frequency} \end{array} \right. \quad (2)$$

and

$$h_{im}(t) = \cos^2(\pi \cdot bw \cdot t) \cdot \sin(2\pi \cdot f_c \cdot t) \quad \left| \begin{array}{l} |t| < \frac{1}{2 \cdot bw} \end{array} \right. \quad (3).$$

The centre frequencies and bandwidths of the filters are chosen according to an approximation for the critical band scale given by Schroeder [8] (Eq. 4):

$$z = 7Bark \cdot \operatorname{arsinh}\left(\frac{f}{650Hz}\right) \quad (z: \text{critical band rate / Bark}) \quad (4).$$

In order to save computational power, the filter implementation is based on a recursive algorithm, which is very fast as compared to straightforward FIR implementations and also somewhat faster than a fast forward convolution. Compared to FFT-like algorithms or multi-rate filter banks, the computational complexity is still very high.

2.1.2. Complex Frequency Smearing

Unlike other models, the frequency-domain smearing is applied prior to the rectification of the filter outputs. The convolution with the masking curve is carried out independently for the filter channels representing the real part of the signal (Eq. 2) and the channels representing the imaginary part (Eq. 3). This preserves the time-frequency resolution and, as the spectral resolution decreases, the temporal resolution increases. The resulting pulse responses of the filters correspond exactly to their new spectral characteristics. Thus, the frequency smearing can be considered as a part of the filter bank.

The steepness s of the upper slope of the masking curve can optionally be chosen level independent according to Brandenburg et al. [6] (Eq. 5)

$$s = -\frac{24dB}{1.0Bark + 0.2 \cdot (z - z_0)} \quad (5)$$

or level dependent according to Terhardt [9] (Eq. 6).

$$s = \min\left(0, -24\frac{dB}{Bark} + \frac{L}{5Bark}\right) \quad (L: \text{sound pressure level / dB}) \quad (6)$$

The steepness of the lower slope of the masking curve is set to 31 dB/Bark. Optionally, a masking curve with a rounded top („rounded exponentials“ according to Patterson [7]) can be realised by applying the same spreading function twice. However, this did not yield any improvement in the prediction of subjectively perceived audio quality.

As the frequency smearing influences the pulse responses of the filters, the level dependent spreading function also results in level dependent pulse responses. The response to a single pulse with high intensity decreases much faster than the response to a pulse of low intensity (*Fig. 4*). This corresponds to the known level dependency of post-masking (e.g. Moore [11]). In the current model this effect can not be used to model level dependencies of time domain masking because the pulse responses are symmetric and they are much too short to model post masking. It would be an interesting task for future research to implement this model with a larger number of more narrow banded filters and introduce a non-linear phase to the spreading function in order to integrate simultaneous and time domain masking into one step of auditory processing.

2.1.3. Adding of Internal Noise

After rectifying the filter outputs by computing the squared absolute values, a frequency dependent offset is added to each filter output (Eq. 7, from Terhardt [9]). This models the influence of internal noise that is assumed to be responsible for the rise of the absolute threshold at low frequencies.

$$E(f_c, t) = E(f, t) + 10^{0.364 \left(\frac{f_c}{\text{kHz}} \right)^{-0.8}} \quad (7)$$

2.1.4. Time-Domain Smearing

In order to model the limited temporal resolution of the human ear, the rectified output signals of each filter are smeared out in time. This time-domain smearing is carried out in two steps. In order to model pre- and post-masking, the signals are averaged over a symmetric time window. The window has a squared cosine shape, which allows a comparison between this model and FFT-based models when the window length is chosen to 1024 or 2048 samples. We found an optimum window length of 400 samples or 8 ms. In order to model the slower decrease of post-masking as compared to pre-masking, a first order low-pass filter is applied afterwards. The time constants depend on the centre frequencies of the filters (Eq. 8).

$$\tau = 0.008 \text{sec} + \frac{100 \text{Hz}}{f_c} \cdot 0.050 \text{sec} \quad (8)$$

2.1.5. Level and Pattern Adaptation

Linear distortions, which may be caused by an unbalanced frequency response of the codec, are less annoying than additive noise or non-linear distortions. The same applies to slow gain fluctuations and constant level differences. In order to separate this kind of artefacts from other distortions, the model aligns the levels of test and reference signal and compensates for the frequency response of the test object. As it is very hard to decide whether a change in the spectral distribution of a signal is caused by additive noise or by an unbalanced frequency response, we had to use some simplified decision criteria. The first assumption we made is, that if the frequency response within a short period differs considerably from the frequency response within a longer period it is very likely that the change in the spectral distribution is caused by additive distortions. The other assumption is, that in the case of an unbalanced frequency response the level difference between input and output will only change slowly between neighbouring filter channels whereas in the case of additive noise the level difference may change rapidly. These assumptions lead to the following algorithm: in the first step, the correction factors for each filter channel are determined by smoothing the signal intensities at the filter outputs by first order low-pass filters and calculating the ratio between test and reference signal. The time constants of these filters are about five times larger than the one used in the time-domain smearing. In the second step, the correction factors are averaged over one critical bandwidth. Afterwards, each filter output is weighted with its corresponding correction factor. The amount of slow gain fluctuations and linear distortions is estimated from the changes of the correction factors over time and frequency.

2.2. Output Values

The time-pitch representations of reference and test signal, which now can be interpreted as excitation patterns, are compared in several ways in order to estimate the amount of audible distortions.

2.2.1. Reduced Noise Loudness

For the estimation of the loudness of the coding noise in the presence of the masking reference signal we tried several approaches. For instance, we tried the noise loudness as proposed by Schroeder et al. [8] and the compressed loudness difference as proposed by Beerends et al. [2]. Additionally, we modified the loudness formula given by Zwicker [10] in order to include the reduction of the coding noise by the masking reference signal. These modifications approximate the loudness of the coding noise when no masker is present ($E_{Ref} = 0$) and depend on the ration between coding noise and masker when the coding noise is much smaller than the masker ($E_{Test} \approx E_{Ref}$). The modification of the Zwicker formula that yielded the best results is given in Eq. 9.

$$NL(f_c, t) = \left(\frac{1}{s_{test}} \cdot \frac{E_{HS}}{E_0} \right)^{0.23} \cdot \left[\left(1 + \frac{\max(s_{test} \cdot E_{test} - s_{ref} \cdot E_{ref}, 0)}{E_{HS} + \beta \cdot s_{ref} \cdot E_{ref}} \right)^{0.23} - 1 \right] \quad (9)$$

In this equation, E_{HS} represents the internal noise according to Eq. 7 and β determines the reduction of the coding noise by the masking reference signal. The values of s_{test} and s_{ref} replace the masking index used by Zwicker [10]. They are calculated from the degree of modulation of the temporal envelope within the corresponding filter channel. For β a simple exponential is used (Eq. 10).

$$\beta = \exp\left(-\frac{E_{test} - E_{ref}}{E_{ref}}\right) \quad (10)$$

2.2.2. Excitation Ratio

A simplified quality measure that can also be used below the threshold of audibility (whereas loudness is only defined above threshold) is the average ratio between the excitations of coding noise and reference signal weighted with the masking index s (Eq. 11).

$$R_{Exc} = \sum_{f,t} s \cdot \frac{E_{test} - E_{ref}}{E_{ref}} \quad (11)$$

2.2.3. Envelope Deviation

The temporal structures of reference and test signal are compared by calculating a sliding cross correlation between the temporal envelopes within each auditory filter (Eq. 12).

$$D_{Env} = \sum_{f,t} 1 - \frac{\int_{-\infty}^t e^{-\frac{T-t}{\tau}} \cdot E_{Ref} \cdot E_{Test} dT}{\sqrt{\int_{-\infty}^t e^{-\frac{T-t}{\tau}} \cdot E_{Ref} \cdot E_{Ref} dT \cdot \int_{-\infty}^t e^{-\frac{T-t}{\tau}} \cdot E_{Test} \cdot E_{Test} dT}} \quad (12)$$

The time constant τ for this sliding cross correlation must be large compared to the time constant used in the calculation of the excitation patterns. Currently we use a value of 250 ms.

2.2.4. Loudness Pattern Deviation

The excitation patterns are transformed into specific loudness patterns using the loudness formula given by Zwicker [10]. From the cross correlation between the specific loudness patterns of reference and test signal, an estimate for the loudness pattern deviation is calculated.

2.2.5. Gain Fluctuation

If the level and pattern adaptation is active and compensates for a changing level difference between reference and test signal, the average temporal derivation of the correction factors is taken as an estimate for the amount of slow gain fluctuations of the test object.

2.2.6. Linear Distortions

If the level and pattern adaptation compensates also for a difference between the spectral envelopes of reference and test signal, the spread of the correction factors between the filter channels is taken as an estimate for the amount of linear distortions introduced by the test object.

Within the available listening test data, the number of test items, where gain fluctuations and linear distortions seemed to have a significant influence on the listening test results, was too low to establish a mapping between these parameters and the basic audio quality.

2.2.7. Other Parameters

Simplified procedures for the calculation of binaural masking level differences and sound localisation have been implemented but up to now yielded no significant improvements.

Besides the reduced noise loudness we also tried to calculate equivalents to other well-known psychoacoustic parameters like roughness and sharpness of the distortion. Both did not seem to improve the performance of the model. Especially the calculation of roughness is a rather complex procedure and there are no examples in literature how to calculate the roughness of one signal in the presence of a masker. Therefore we are not able to decide whether roughness does not contribute much to the basic audio quality or whether the simple roughness models we have used were not sufficient for this task. It is also possible that roughness is already sufficiently included in the calculation of the reduced noise loudness because both the masking index s in Eq. 9 and the roughness are calculated from the degree of modulation within each auditory filter.

3. Results

3.1. Validation of the Model

We checked the model against the subjective results of several listening tests performed by ITU and MPEG during the last six years. As the relation between the output values of the model and the „5 Grade Impairment Scale“ used in the listening tests is normally not linear, we applied a sigmoidal mapping function to our model values before comparing them to the listening test results (shown as a solid line in *Fig. 5* to *Fig. 8*). The standard error s and cross correlation coefficient r between mapped model values and listening test results were calculated for several sets of listening test data. Especially the „reduced noise loudness“ showed rather high correlations with the subjective data. In *Fig. 5* to *Fig. 8* the relation between the „reduced noise loudness“ and the listening test results of the ISO/MPEG 1990 test (headphones only) and the ITU 1993 tests is shown. We plotted the quality differences on the „five grade impairment scale“ against a logarithmic representation of the „reduced noise loudness“. For most of the available sets of listening test data, the correlations between this parameter and the subjective data were comparable to the best published results of other perceptual models (e.g. [2], [3]).

3.2. Influence of Temporal and Spectral Resolution

The influence of the spectral and temporal resolution on the correlations between model values and listening test results was smaller than expected. Increasing the number of filter channels from currently 80 to 120 or more did not yield any improvement. Decreasing the spectral resolution down to 50 or 30 filter channels affected the correlations for some sets of listening test data but in most cases the results did not change significantly.

Varying the length of the symmetric time window used for pre- and post-masking, we found that the prediction errors increase rapidly when the time window gets shorter than the optimum of approximately 8 ms. When the time window becomes longer than the optimum, the prediction errors increase much more slowly. Varying the time constant within the first order low-pass used in the time-domain smearing, we found similar results. Time constants larger than 20 ms yielded a moderate increase of the prediction errors whereas the prediction errors increased very fast when time constants below 8 ms were used. The dependence of the time constants from the centre frequencies of the auditory filters played only a secondary role. Leaving away this dependency did not have any significant influence on the prediction errors.

3.3. Influence of the Spreading Function

The level dependent spreading function according to (Eq. 6) and the level independent spreading function according to (Eq. 5) yielded about the same results when correlating the model output values with the listening test data. Level independent spreading functions with a simple exponential decrease of the upper slope yielded slightly lower correlations when the slope rate was between 10 and 20 dB/Bark and rather poor correlations when the slope rate was smaller than 6 or larger than 24 dB/Bark. Using a rounded exponential shape according to Patterson [7] also yielded significantly poorer correlations as compared to the simple triangular shape of the spreading function.

3.4. Performance of Different Output Parameters

The parameter that yielded the highest correlations between model predictions and listening test data was the reduced noise loudness according to Eq. 9. The noise loudness as proposed by Schroeder et al. [8] gives rather high prediction errors in the context of our model. For the compressed loudness difference as proposed by Beerends et al. [2] we found about the same optimum compression ratio as given in [2] but much larger prediction errors. This is probably due to the different calculation of the excitation patterns between the models. The excitation ratio yields always slightly larger prediction errors as compared to the reduced noise loudness but the correlations are still rather high. Envelope deviation and loudness deviation did not yield such high correlations but the results are still comparable to the performance of other perceptual models.

4. Problems

Even though in the context of our model Eq. 9 turned out to be a very good approximation for the reduction of the noise loudness by the masking reference signal, it is by far not optimal. There is no easy way to handle the case when the expression $s_{test} \cdot E_{test} - s_{ref} \cdot E_{ref}$ gets smaller than zero. Fortunately, this happens only in very few cases. We are currently trying to solve this problem but up till now all solutions to this problem increased the prediction errors for the total of the available databases.

The level and pattern adaptation may cause problems because under some conditions it might fail to distinguish between additive noise and linear distortions. It can happen that additive noise is mistaken for a linear distortion and gets partially suppressed. In cases where linear distortions or gain fluctuations are responsible for the perceived audio quality the current model will get problems because there are too few examples for this case within the available sets of listening test data. Normally, when linear distortions occurred, the remaining distortions identified as additive noise were sufficient to explain the listening test results. Therefore, we could not determine a mapping between linear distortions and the basic audio quality.

Some examples for coders where our model could not produce satisfactory predictions were the NICAM codec in the MPEG 1991 test, the RAI codec in the MPEG 1994 NBC test and the AWARE codec in the ITU 1992 test. In the case of the NICAM codec we could already identify the reasons why our model had problems. It is mainly because of the drawbacks we mentioned above and we think that these problems will be solved very soon. In the case of the RAI codec and the AWARE codec we still do not know the reason for the poor predictions.

5. Future Research

Some improvements of our model can be expected from implementing refined models for the binaural processing. Furthermore, we are still trying to improve the calculation of the reduced noise loudness because of the problems mentioned above. A more reliable distinction between linear distortions and additive noise could be achieved by including the envelope correlation between reference and test signal in the pattern adaptation algorithm. Some work will be spent in order to combine different output parameters into one quality measure. Up till

now, we only tried linear or polynomial combinations, which did not yield significantly better correlations as compared to the reduced noise loudness alone.

6. References

- [1] Brandenburg, K. H.; Sporer, Th.: NMR and Masking Flag: Evaluation of Quality Using Perceptual Criteria. Proceedings of the AES 11th International Conference, Portland, Oregon, USA, 1992, S. 169-179.
- [2] Beerends, J. G.; Stemerdink, J. A.: A Perceptual Audio Quality Measure Based on a Psychoacoustic Sound Representation. *J. AES*, Vol. 40(1992), No. 12, December, S. 963-978.
- [3] Beerends, J. G.: Measuring the Quality of Speech and Music Codecs, an Integrated Psychoacoustic approach. Contribution to the 98th AES Convention, Paris, February 1995, Preprint 3945.
- [4] Paillard, B.; Mabillean, P.; Morissette, S.; Soumagne, J.: PERCEVAL: Perceptual Evaluation of the Quality of Audio Signals. *J. AES.*, Vol. 40(1992), No. 1/2, January/February, S. 21-31.
- [5] Colomes, C.; Lever, M.; Dehery, Y. F.: A Perceptual Objective Measurement System (POM) for the Quality Assessment of Perceptual Codecs. Contribution to the 96th AES Convention, Amsterdam, February 1994, Preprint 3801.
- [6] Brandenburg, K. H.: Ein Beitrag zu den Verfahren und der Qualitätsbeurteilung für hochwertige Musikcodierung. Ph. D. Thesis, Erlangen-Nürnberg, 1989.
- [7] Patterson, R. D.: Auditory Filter Shapes Derived with Noise Stimuli. *J. Acoust. Soc. Am.*, Vol. 59, No. 3, March 1976, pp. 640-654.
- [8] Schroeder, M. R.; Atal, B. S.; Hall, J. L.: Optimizing digital speech coders by exploiting masking properties of the human ear. *J. Acoust. Soc. Am.*, Vol. 66(1979), No. 6, December, S. 1647-1652.
- [9] Terhardt, E.: Calculating Virtual Pitch. *Hearing Research*, Vol. 1(1979), S. 155-182.
- [10] Zwicker, E.; Feldtkeller, R.: *Das Ohr als Nachrichtenempfänger*. Stuttgart: Hirzel Verlag, 1967.
- [11] Moore, B. C. J.: *An Introduction to the Psychology of Hearing*. London: Academic Press, 1989.
- [12] ISO/MPEG: MPEG/Audio Test Report, ISO/IEC JTC1/SC2/WG11 (N0030), 1990.
- [13] ITU-R TG 10/2: CCIR listening tests, network verification tests without commentary codecs, final report. Doc. 10-2/43, 1993.
- [14] ITU-R TG 10/2: Listening tests for the assessment of low bit-rate audio coding systems. Doc. 10-2/47, 1993.

7. Figures

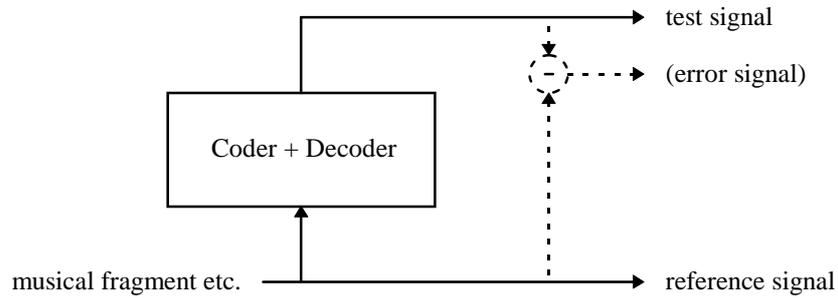
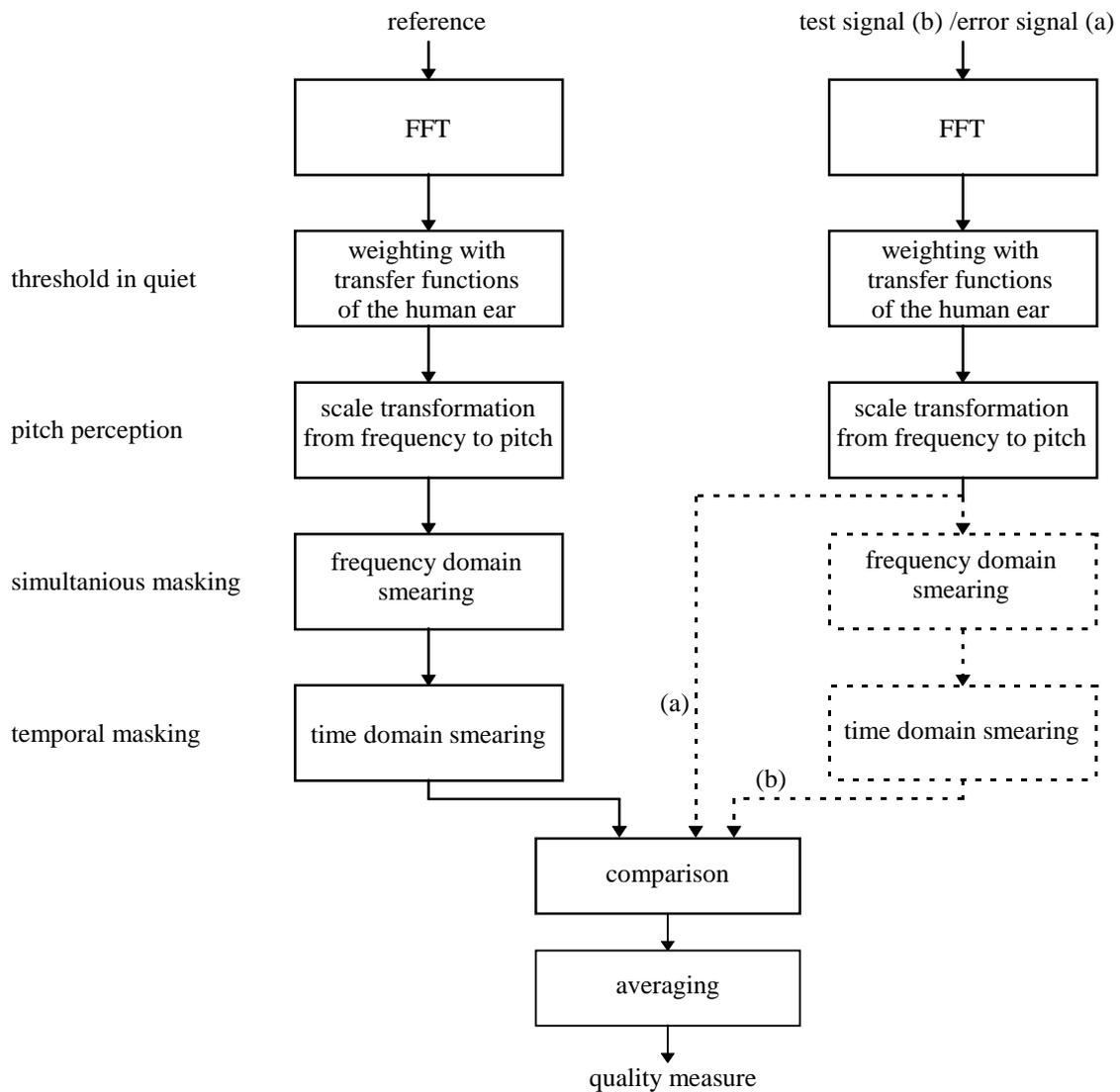


Fig. 1: Input signals for a perceptual measurement system.



*Fig. 2: Scheme of a perceptual measurement system.
(a): NMR, (b): PAQM, PERCEVAL, POM*

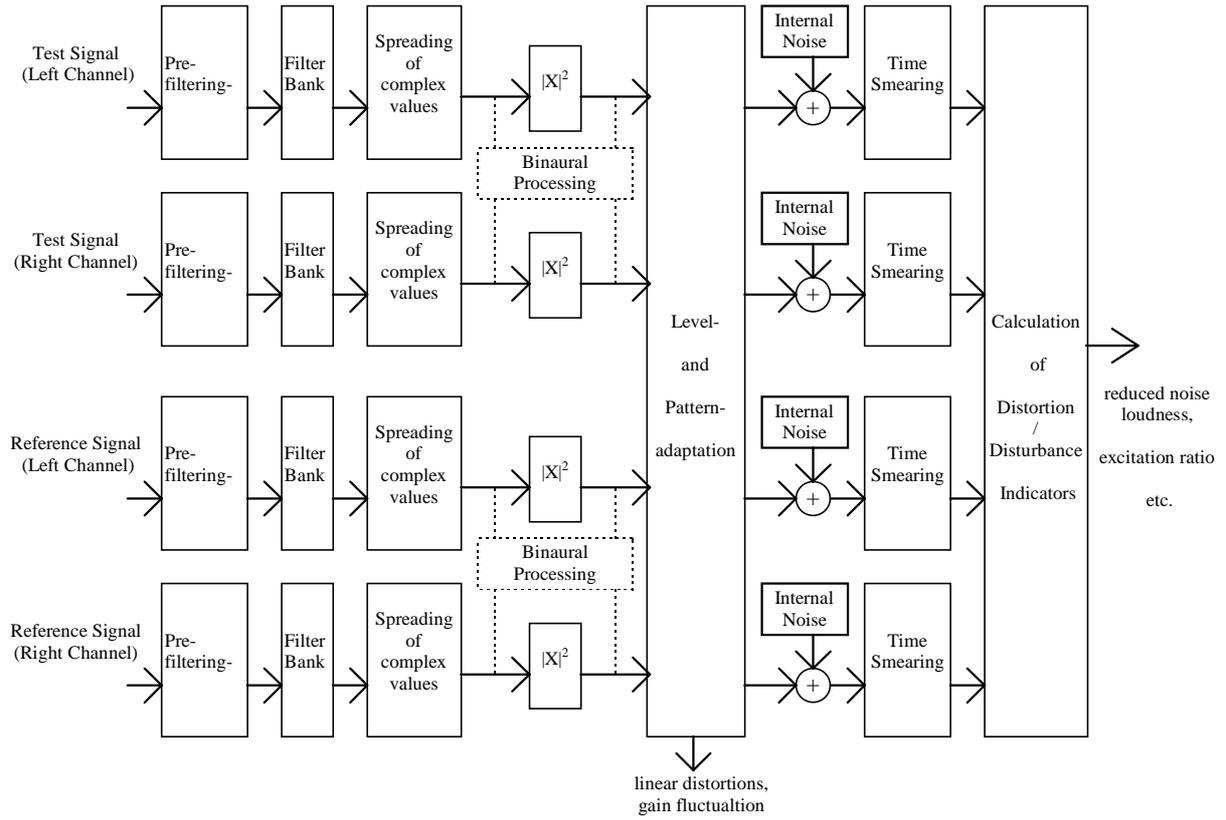


Fig. 3: Structure of the model.

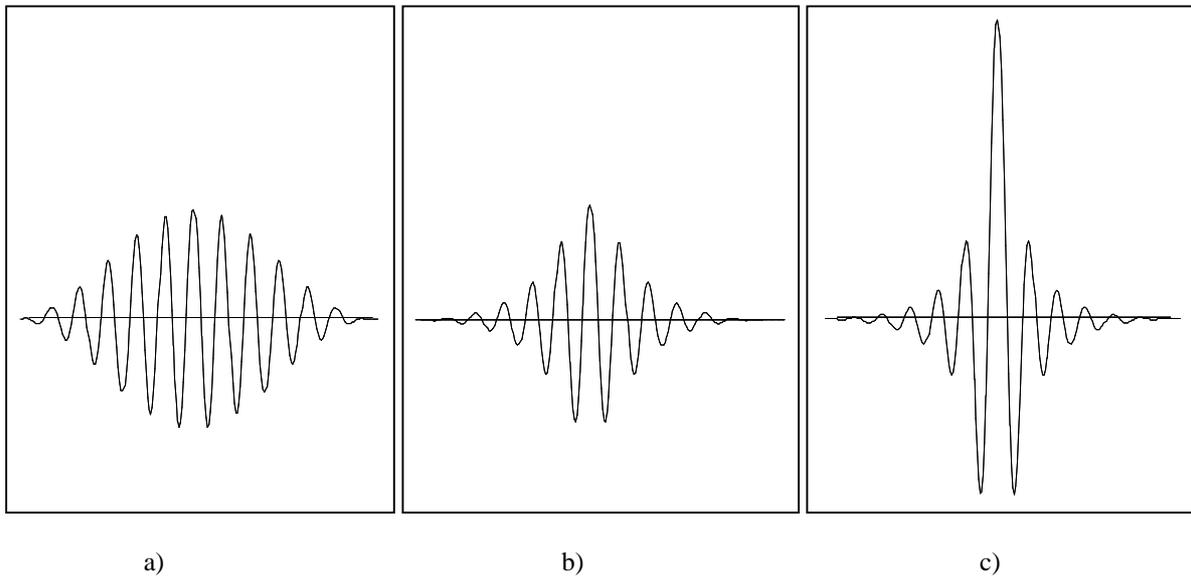


Fig. 4: Pulse responses at one filter channel prior (a) and after (b, c) the complex spreading operation for two pulses of different energie (b: low energie, c: high energie). The shape of the pulse response after the complex spreading operation corresponds nicely to the post-masking curves shown in [10].

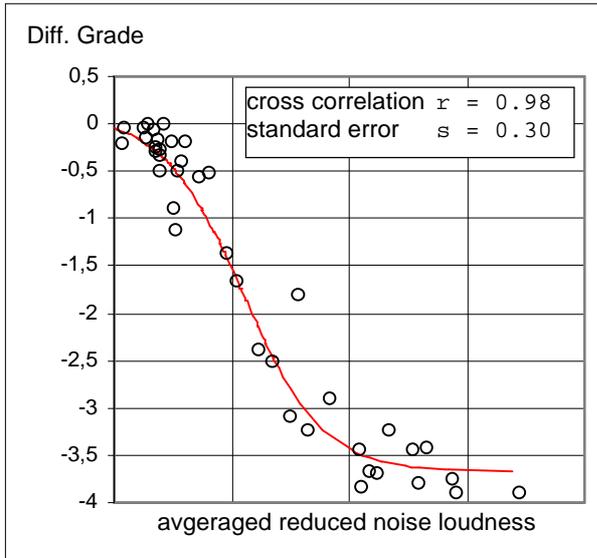


Fig. 5: Results for 40 items (Musicam and SB-ADPCM) of the ISO/MPEG 1990 test (headphone presentation) [12]

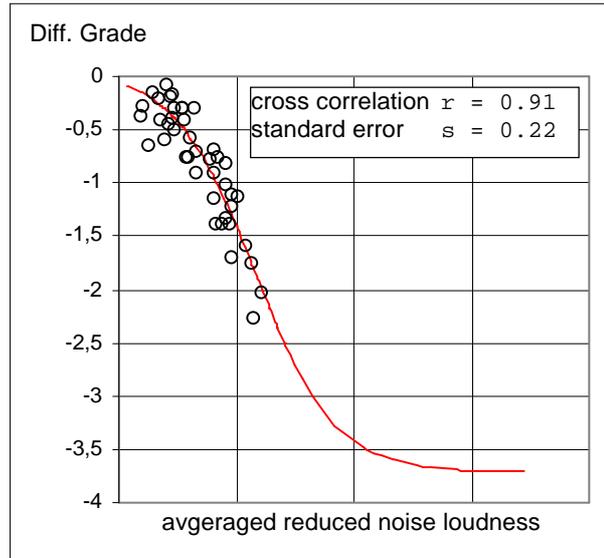


Fig. 6: Results for 42 items (cascaded codecs) of the ITU 1993 test [13]

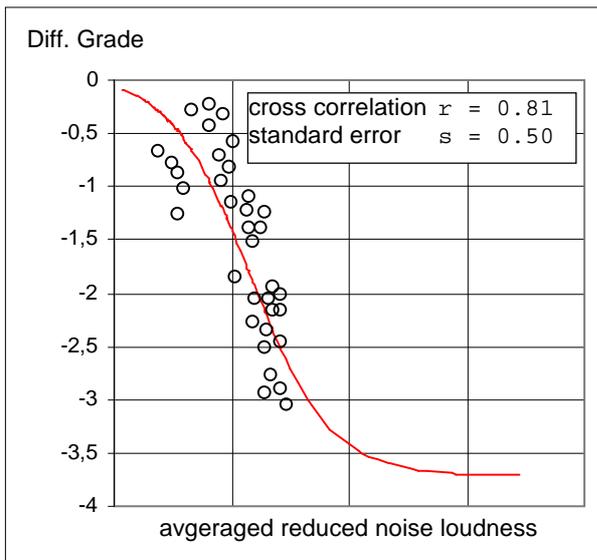


Fig. 7: Results for 35 items (commentary stereo) of the ITU 1993 test [14]

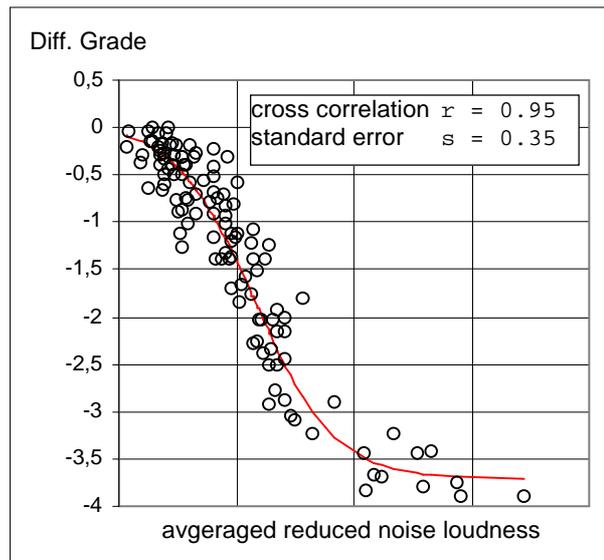


Fig. 8: Results for all three tests put together